Deriving Products from a Multi Resolution Database using Automated Generalisation at Ordnance Survey

Regnauld Nicolas*, Lessware Seb, Wesson Christopher*, Martin Paul****

* Ordnance Survey (GB), Adanac Drive, Southampton, SO16 0AS, United Kingdom
** 1Spatial, Tennyson House, Cambridge Business Park, CB4 0WZ, United Kingdom

Abstract.

Ordnance Survey is currently conducting an investment programme aiming at building and maintaining its future production systems. The architecture built by the system relies on a Multi-Resolution Database to store the data in the form of reusable data components available at different levels of detail. It also relies on a Service Oriented Architecture to build the workflows necessary to deliver highly automated production systems. With the first production system recently completed, this paper describes the architecture designed and implemented by the programme, the technology provided by 1Spatial to perform the automated generalisation, as well as the product created by this system, OS VectorMap® District v1.0. The paper also highlights the benefits of this new version compared to the beta version which was derived using research prototypes.

Keywords: Map production system, Automatic generalization, Workflow, Multi Resolution database, Service Oriented Architecture

1. Introduction

Ordnance Survey has recently completed its 'Geospatial Data Management System' (GDMS) a large programme to upgrade its systems to capture and manage its large scale data. A few years ago, once we had enough confidence about what GDMS would deliver, in terms of richness of data and its structure, Ordnance Survey has started to invest in another programme, called the Multi Resolution Data Programme (MRDP) to exploit this new database.

This paper focuses on OS VectorMap District v1.0, and its production system, the first system which has been designed and built by MRDP to derive a new product automatically from large scale data. OS VectorMap District has seen two releases prior to this 1.0 version. The alpha version was released in April 2010. A beta version was released a year later, after applying a range of enhancements to the product, based on the feedback received. The description of these versions of the product can be found in (Revell et al 2011). It describes the main generalisation algorithms used to derive the product. While both alpha and beta versions were produced using software prototypes and manual processes were involved to trigger all the processes required from the data import to the publication of the product, version 1.0 is produced by a proper enterprise system. (Regnauld et al 2012) gave an overview of the changes that were being introduced by MRDP to build a robust production system for OS VectorMap District. Now that the system is finished, this paper present the system built and the impact it has on the product. The paper is organized as follows: Section 2 explains why MRDP was started and its objectives. Section 3 provides an insight into the system architecture built by MRDP for deriving products. It presents the Multi Resolution Database, the high level service oriented architecture of the system, and the technology provided by 1Spatial to perform the automated generalisation . Section 4 describes OS VectorMap District v1.0, the improvements that have been made to the product since the beta version, and examples of different usages of the product. Section 5 concludes the paper and discusses future improvements.

2. Context and objectives

Ordnance Survey has recently started a large programme of work to change the way products are created. MRDP was triggered by a clear change in customer needs. Feedback received from customers show that the current Ordnance Survey products no longer fully meet their needs. This is highlighted by the fact that customers are far more informed and will now articulate their map data requirements, as opposed to choosing from a fixed set list of products which was the case in the past. The first objective of the programme is therefore to deliver a more flexible production system, to allow the organisation to evolve its product portfolio to keep it in line with ever changing customer requirements. Minimising production costs is also a key objective of the programme. The last main objective is to increase the consistency between products.

At the centre of the architecture proposed by MRDP is a multi-resolution database that provides reusable data components. The architecture is explained in section 3.1. Here we explain how it helps delivering the three main objectives of the programme.

2.1. Making product creation easier

The Multi-resolution database provides us with a library of data components at different levels of detail, that we can use to create new products. Take a hypothetical scenario of creating a new mid-scale cycling map. If we were to create this from scratch then we would have to either (a) use another product as a starting point, or (b) take source data from our large-scale database.

The problem with using another product is that this becomes what we call 'product chaining'. Chaining together products is not recommended because you are rarely going to have the best specification for your new product already existent within another product. By building on a pre-existing product you are compromising the usefulness of the map or at least the clarity of the information to the map user. It also means that the update and refresh of one is always tied to the other.

The second option, involves an enormous amount of work to extract data, filter that data and align its attribution properties to specification, build databases, test and fix errors in the processing and perhaps biggest of all – to generalise the data.

By having a multi-resolution database, we can take scale relevant data that has already been generalised, already been tested and is already stored in a database without product-specific but importantly with product-friendly attribution.

2.2. Increasing product consistency

If the footprint data in all products has come from the same source and the number of different generalisation levels is constrained, then the final map output has both a signature and no sign of any difference of opinion. This leads to greater product consistency in terms of appearance. This commonality in the look and feel of our mapping strengthens our brand and makes our map products instantly recognisable as Ordnance Survey maps. Consistency across products is also ensured by the fact that they share the same creation and update process, so any real world changes are added to all products at the same time, they should never be out of synchronisation beyond the factor of product release dates.

These consistencies lead to a recognisable set of products that now, more than ever, can be used in conjunction with and to complement one another. For products like OS VectorMap, this has allowed us to think about generating product families that are consistent through the scales. This in turn allows us to make better web map services.

2.3. Increasing efficiency

The last main objective of the programme is to provide the infrastructures that allow Ordnance Survey to create and maintain products in a more cost effective way. This means being able to derive many products from a single source of data, with a high level of automation. Advances in research on automated generalisation have made this possible. An overview of the research activities in the domain during the last ten years, as well as examples of implementations in production systems can be found in (Mackaness et al 2007) and (Burghardt et al 2013).

MRDP not only focuses on deriving new products, but also aims to build systems that maintain them in a cost effective manner. The programme therefore is aiming at producing a system that supports incremental updates of its products.

3. System Architecture

3.1. Multi resolution database

At the heart of the architecture developed by MRDP lies a multi-resolution database. The idea behind it is very similar to Multi-Representation databases, which have been widely studied and are based on the DLM/DCM (Digital Landscape/Cartographic Models) principles, first presented in (Grunreich 1985) and now widely adopted (Balley et al 2004, Trevisan 2004, Bobzien et al 2007). We have used resolution instead of representation to avoid confusion with cartographic representation. The core of the database is made of the different content databases representing the world at different levels of detail, without symbolisation constraints, which are only introduced in product specific databases.

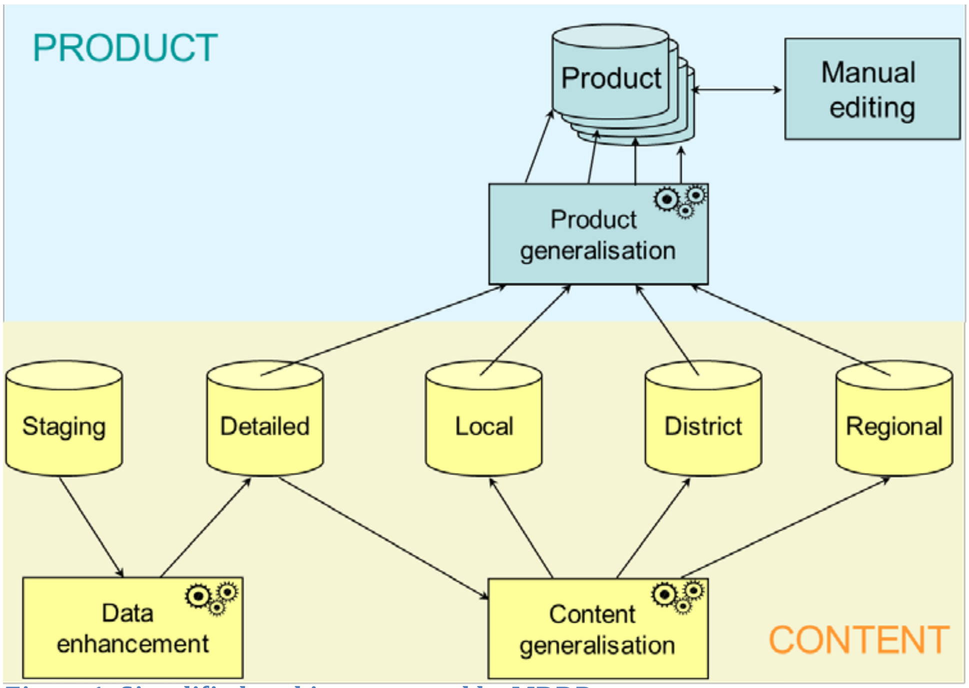

Figure 1: Simplified architecture used by MRDP

For the MRDP systems, the source data come from the staging database. This is a physical copy of the GDMS maintenance database which is continuously updated. MRDP systems do therefore not interfere with all the systems in use to update the maintenance database. The first type of process that may be applied regroups processes that perform data enrichment, i.e. processes that make explicit information which is implicitly present in the data, like deriving urban extents. This is stored in the detailed content database. This is then the source for the derivation of all the other content databases, through content generalisation processes. Each content database is then used as a source for a family of products requiring data at this resolution. Some product specific generalisation is performed at this stage. Products databases store the data that directly support the publication of a specific product. These databases can be manually edited, which is usually required for rich cartographic products. This is illustrated in Figure 1.

3.2. High level architecture

MRDP is implemented according to the Service Oriented Architecture (SOA) methodology. With SOA, individual components known as services are orchestrated together in a workflow to act as in bigger, more complex system. Individual services are designed to perform discrete tasks and have

a well-defined set of operations. One of the benefits of SOA is that it encourages software reuse where a service offering a particular function can be called in many places in the workflow and interoperate with other services in different ways.

Within a SOA framework a service must be designed according to the following rules:

- A service must perform one task and one task only.
- A service must always complete its operation, and inform the calling application on its successful completion or failure.
- A service whose operation has failed should report back to the calling application why it failed.

By following these basic rules a SOA based framework gives us a very stable platform under which to operate a complex system such as MRDP. We guarantee that at any point in the workflow we can tell whether a service has completed successfully or not and we can build business logic into the workflow tier to handle and manage any problems that occur.

MRDP is composed of a number of enterprise applications that perform a particular role in the MRDP workflow. By using a SOA framework we are able to effectively join these different technologies together by pushing the orchestration business logic into the workflow tier. The means a workflow component is able to combine technologies which have not been designed to work together.

The workflow technology used in MRDP is the Oracle Workflow Manager (OWM). Figure 2 shows at a very high level the major service components

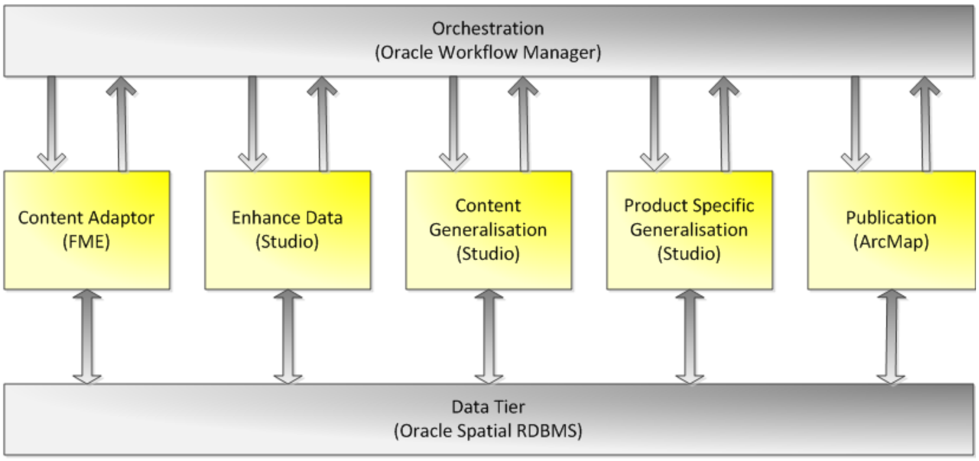

Figure 2: MRDP high level workflow

that form the MRDP workflow. It combines the four main geospatial technologies used in MRDP - SAFE Software's FME, 1Spatial's Radius Studio, ESRI's ArcMap and Oracle Spatial. The OWM workflow layer at the top of the figure allows these three technologies to interoperate in a single system. As each service performs an operation, the data it works on is passed around in the Oracle Spatial RDBMS data tier.

The other important aspect of the MRDP Architecture is that it allows us to scale up the system through adding additional servers. System scalability is of fundamental importance to MRDP. As we add new content and new products to the portfolio the MRDP system must be capable of scaling up to support that demand. This is in part allowed through the flexibility of the SOA framework – however our choice of technologies has also helped with system scaling, particularly around the data enhancement and generalisation capabilities.

The FME and ESRI technologies are implemented as REST-ful services. Those technologies do not scale in their own right; however by hosting those technologies across a farm of servers we can use the SOA Framework to distribute requests across them, effectively allowing us to scale almost linearly.

In contrast, 1Spatial's Radius Studio technology is an enterprise solution that has from the outset be designed to scale. Again this is physically implemented by hosting a number of instances of Studio on a farm of servers, but in the case of Radius Studio there is only a single entry point to that

service. Therefore the SOA workflow only needs to be aware of the single entry point and the system scaling is left to the Studio application itself. This achieves the same near linear scaling as for the FME and ESRI solutions, however in this case the SOA Framework is not responsible for distributing the requests.

In both instances the system scalability inherent in MRDP has achieved a significant improvement in performance when generating the latest refresh of the OS VectorMap District product (v1.0).

3.3. Key technology used

As a result of a competitive tender process, Ordnance Survey selected 1Spatial to provide a platform for building an automated generalisation process. This platform makes use of the components Radius Studio and Radius Clarity (now rebranded as 1Integrate and 1Generalise).

These components are used to perform the data enhancement, content generalisation and product generalisation tasks as presented in the next section and in Figure 1.

3.3.1. Radius Studio

Radius Studio (now rebranded as 1Integrate) is a rules-based component that allows data quality rules and data processing actions to be defined and applied to spatial data. Rules are defined as a tree of spatial or non-spatial first-order logic predicates and processing actions are defined as a sequence of predicates and geometric processing operations, making the platform well suited to content (model) generalisation tasks. Radius Studio can be operated as a standalone application or can be deployed as a SOAP web service within an enterprise architecture.

Radius Studio processes are deployed as J2EE components over a grid of servers to provide high availability and linear scalability. The grid technology is implemented using the OpenSource (GPL) GridGain framework, allowing many processing engine nodes to be deployed across a network where each processing job can be handled in parallel by a different node. It is currently in operational use by Ordnance Survey as a key component of their GDMS. The GDMS dataset contains over 500 million individual features and the system needed to be capable of being published continuously by enforcing automated data quality checks and running data publication processes.

Data is read from an enterprise database such as Oracle Spatial and, like at Ordnance Survey, can optionally be managed using Oracle Workspace Manger to provide long transactions during data maintenance.

Radius Studio rules and processing actions are authored using a web-based user interface that does not require programming skills and are managed as XML in a rules repository. The list of processing functions used within the rules can be extended with customer-specific operations that are written in Java and deployed into the processing engine.

3.3.2. Radius Clarity

Radius Clarity is a desktop application that uses 'Agent' technology to make map objects 'self and context aware'. This allows objects to 'co-operate' to achieve acceptable automatic cartographic (product) generalised results by trying a number of strategies to achieve local feature goals and selecting the set of results that provide the 'happiest' global result across the entire dataset. This approach has been shown to provide production quality results for automatic cartographic generalisation (Duchêne, 2003).

Radius Clarity provides a graphical user interface that allows the process to be configured in detail and then tested and debugged. As well as configuring the processes, additional algorithms and capabilities can be written in Java and plugged into the framework. The processing is invoked via a user interface or from the command line.

3.3.3. Radius Clarity / Radius Studio integration

Although both use the same underlying object oriented technology and geometry engine, these two components have very different strengths and capabilities: Radius Studio for content (model) generalisation with enterprise scalability and Radius Clarity for Agent-based cartographic (product) generalisation. They both make use of a flexible object model that can be used to implement any spatial domain models and contain a versioned object technology with very low query latency supporting efficient spatial processing.

The integration of Radius Clarity's Agent technology with Radius Studio for the MRDP programme was achieved by building Radius Clarity's Agent libraries into the Radius Studio engine and providing new built-in functions for invoking these Agent processes from within Radius Studio to achieve cartographic generalisation. This integration means that Radius Clarity's approach to automated cartographic generalisation became available for deployment within the scalable, enterprise service-based architectures and Radius Studio could therefore now be used to perform end-to-end automated map generalisation within a National Mapping Agency production environment. For MRDP, Radius Studio's web service API is invoked automatically from a Business Process Execution Language (BPEL) workflow during the content and product stages of the architecture.

The ability to combine model and cartographic generalisation as a set of business rules within a single scalable environment opens significant possibilities for the future, not least the ability to ensure that spatial databases are maintained in a constantly product ready state for on demand mapping and digital product generation purposes.

4. Overview of OS VectorMap District v1.0

A key element of Ordnance Survey's product strategy is a suite of map products that can be customised to create bespoke contextual maps for their websites and applications, generated from the large scale database which our data collection activities update through GDMS with around 5000 changes per day. The OS VectorMap product family is the outcome of this and currently includes the street-level OS VectorMap Local and a district-level equivalent, OS VectorMap District. Future OS VectorMap products will cover other resolutions, giving customers a consistent zoom-stack of mapping from local through to national level.

OS VectorMap District is therefore a district-level contextual map product built from the multi-resolution database and developed from an award-winning beta (Avenza Award 2011 for Electronic Mapping, from The British Cartographic Society). It is available as a part of OS OpenData and current users include local authorities, emergency services, and the insurance industry. It is also an ideal entry-level product for market demographic displays and can be used by anyone for sharing statistics or for neo-cartography.

By district we intend to offer a category of scale somewhere between what is traditionally referred to as the large scales and the mid scales. However to offer an indicator, the product is produced at a reference scale of 1:25 000, and our cartographic designers tend to prefer to view the raster on-screen at around 3.5 metres per pixel (m/px).

The product consists of the following features many of which are separated into sub-features and have some supplementary attribution: Administrative boundaries, airports, buildings, electricity transmission lines, foreshore, glasshouses, heritage sites, land, motorway junctions, named places, ornament, public amenities, railway stations, railway track, railway tunnels, roads, road tunnels, roundabouts, spot heights, surface water (as areas and as lines), tidal boundaries, tidal water and woodland.

4.1. Improvement since beta version

The changes made to the product and its production system fall into three main categories: changes that have made the product easier to use, changes that have increased the quality of the product, and changes that have improved the production system.

4.1.1. Making the product easier to use

The following improvements have been made to improve the usability of the product:

- A full colour raster product has been made available in addition to updated backdrop raster. The full colour style is intended to be a complete map that works across all screen types and digital printers to provide context to geographic information. The backdrop style is intended to provide a base map for customers who wish to overlay their own geographic data onto the map. Both styles offer customers map hierarchy without using prime or pure colours, so even the full colour style will facilitate the addition of logos.
- The product is now available in GML 3.2 format, in addition to shape files. This should work easily with many GIS, without the need to purchase additional software.
- Raster versions of the product are now in GeoTIFF format, so customers no longer need to download separate georeferencing files.
- OS VectorMap District v1.0 makes use of the latest version of Ordnance Survey's improved stylesheets. This includes for the vector product, layer files (.lyr) and styled layer descriptor files (.sld) and for the raster GeoTIFF (.tif) files, all of which are based upon the same stylesheets and are offered in two variants, full colour and backdrop styles.
- Feature codes have been added to make it easier for some customers (e.g. CadCorp users) to style the data.

4.1.2. Improving the quality of the product

The following improvements have been made to improve the quality of the product:

- Railway: sidings have been more consistently removed. On the raster product, the depiction at railway bridges (rail on top of road) has been improved. Specific errors omissions from the beta version have also been fixed.

- Railway tunnel identification and extents have been improved in our data, which allowed our cartographic designers to improve the tunnel peck (dashed line) depiction.
- Road: Dual carriageways have been collapsed into a single centreline (see Thom 2005 for information about the automated process used). Small roundabouts have also been collapsed to single points. This has enabled us to provide a much better depiction of the roads on the raster product, and the vector product will be easier to style.
- Pontoons and jetties have been removed from inland water.
- Isolated buildings at the end of jetties and piers have been removed from the sea.
- Text has been improved on the raster and will soon be applied to the vector. In the raster, Corbel has been replaced with Camphor Pro, which is a crisper, easier font to read, especially on screen. The new colour palette has been assigned to improved text groupings, e.g. now all communications text is in the same colour but is a different black-grey from the rest of the text. Overall the selection of text has been improved, and the text placement has benefited from the use of a later version of ArcGIS and Maplex.

4.1.3. Improving the production system

The most important change for this release is that a proper production system has been built to derive the product from the large scale data. Here are the key facts and benefits about this new system:

- The production system is now based on Oracle Spatial databases rather than the tiled shape files used as input for the beta version.
- The data now follow the MRDP architecture, making the data more usable by separating it out into real-world feature types in Content Staging, a separation that continues through the content stores.
- Buildings generalisation process is now very robust and not prone to random failures.
- Processing has been made scalable using Radius Studio servers, and consequently runs much faster than beta.
- Repeatable processes have been developed to publish shape, GML and Raster Tiffs.
- Automated quality checks have now been written to trap major errors.

- Data quality issues have been identified and work is underway to correct these in GDMS.

4.2. Product examples, and examples of use

Two stylesheets have been released with the product: Full colour style and backdrop style, they can be seen in Figure 3 and 4 below.

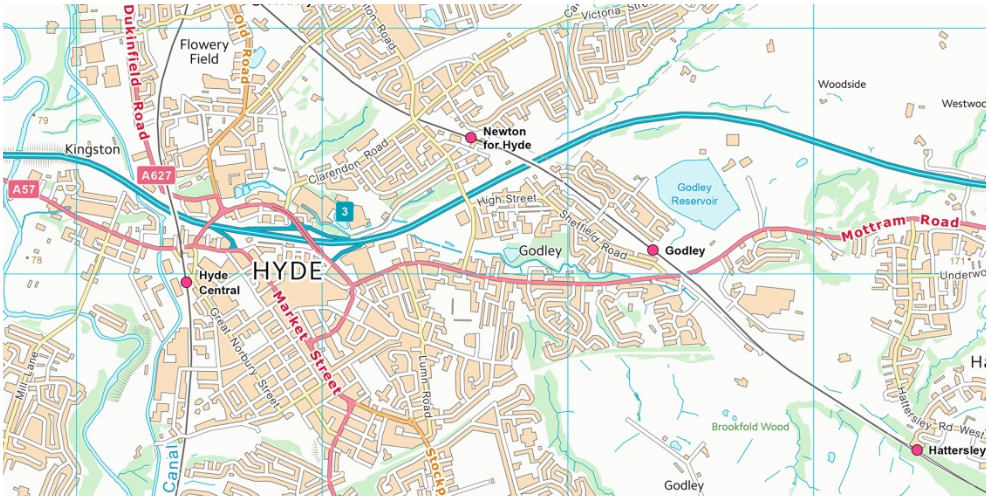

Figure 3: OS VectorMap District - Full colour style

Figure 4: OS VectorMap District - Backdrop style

The vector product can be loaded into a GIS and used as a map product or a set of footprint polygons on which to attach third party geodata. The different features and their attribution allow the user to filter to their own requirements and the data can easily be restyled.

In Figure 5, the product has been loaded into ESRI ArcGIS using the full colour layer file available from the Ordnance Survey website. Any further data loaded into the system will then have geographic context and can exploit the underlying attribution.

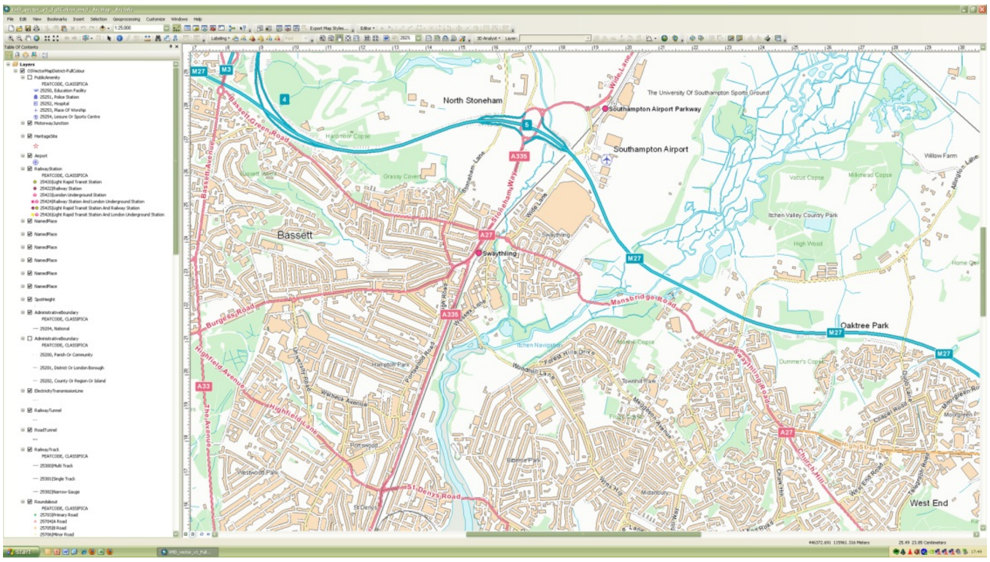

Figure 5: OS VectorMap District shapefiles with the layer file applied as a contextual map within a GIS

In Figure 6 third party OpenData from the Office of National Statistics (ONS) showing a number of fires in localities has been joined to the OS VectorMap District building polygons which have been colour-coded accordingly. From the same third party data, the number of casualties in a given year (2006) has also been indicated using the ONS polygons. Finally the location of five fire stations from other Ordnance Survey data has been shown as point symbols.

Figure 7 shows how not all of the features in OS VectorMap are necessarily required. It also shows how a cartographer might decide to completely transform the appearance of the product by restyling and adding another dataset such as a height product.

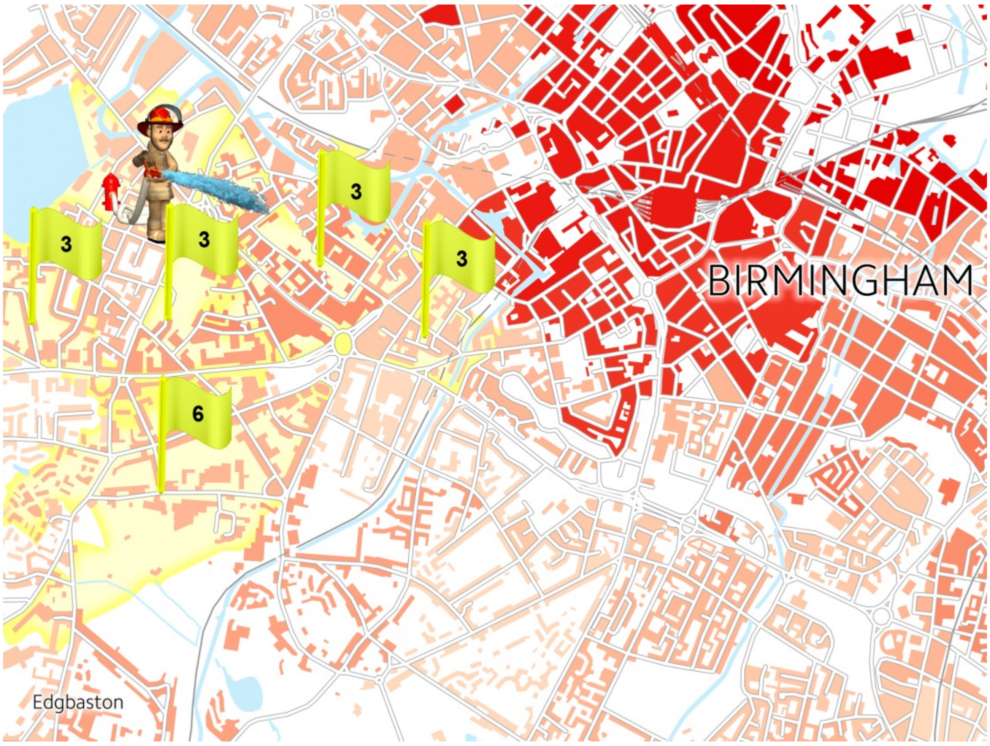

Figure 6: A map to show fire risk in Birmingham

Figure 7: OS VectorMap District combined with a height product to show flood risk in Carlisle

In Figure 8, the GCSE performance of schools across Ipswich has been shown graphically using pie charts. As the schools are obviously features with a location and this location is likely to be a factor in the statistics, it makes sense to apply some geographic context. Here the backdrop style of OS VectorMap raster has been placed beneath the pie charts which are positioned at their relative school's location.

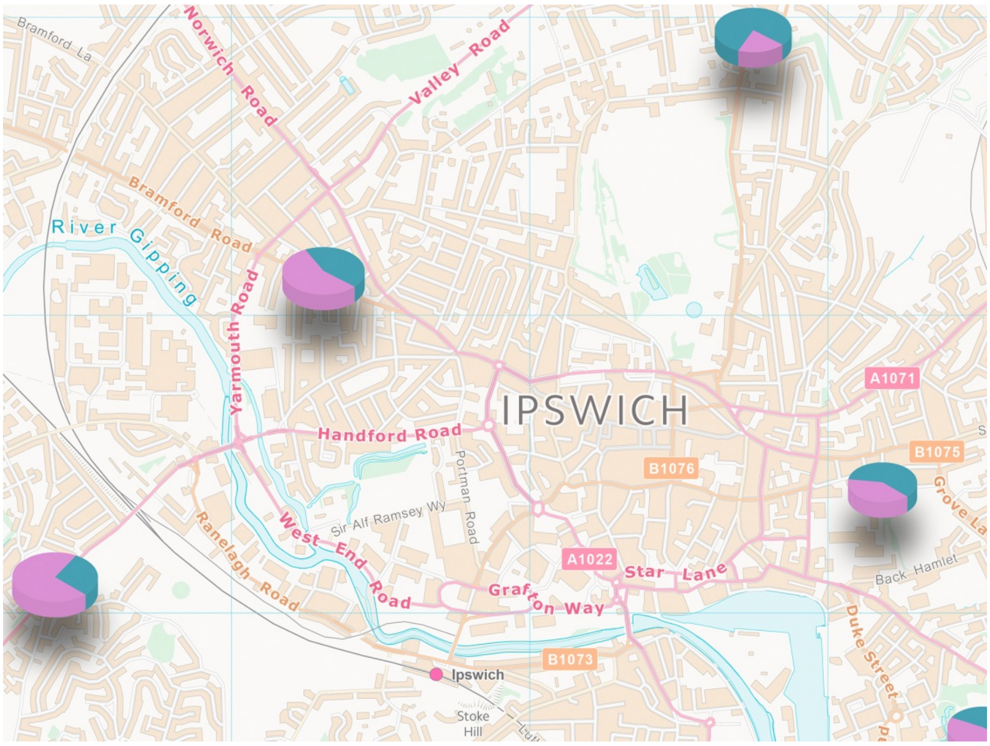

Figure 8: OS VectorMap District - Backdrop style raster adding geographic context to school performance statistics

5. Conclusion

We have described in this paper the reasons why Ordnance Survey is investing in MRDP, and how the programme has delivered its first automated system to derive a product from large scale data.

MRDP has been designed with efficiency in mind. The Multi Resolution Database provides reusable data components, and the system architecture provides a solid framework for building more production system allowing the easy reuse of software components already developed for others.

The next step for MRDP is to develop production systems for other products. However, a number of components still need to be completed. The

most important are the incremental updating capabilities, and the integration of the manual editing platform. As OS VectorMap District v1.0 does not require manual editing, and can be regenerated entirely from GDMS data within a few weeks, these two aspects were low priority. They will become high priority when MRDP targets products with richer content, that require some level of manual finishing.

In addition to improving its own systems, MRDP also provides valuable feedback about the base data, in particular the types of improvement that would allow the programme to improve the quality of the derived products. For example, the way names are classified in our database could be improved, which would allow the automatic process to draw them in the right colour and style. Named extents could also be improved, which would allow the system to more accurately position the text and for text size to be more reflective of the feature's geographic area.

References

Balley, S.; Parent, Christine; Spaccapietra, Stefano, 2004: Modelling geographic data with multiple representations. *International Journal of Geographical Information Science*, vol. 18, num. 4, p. 329-354

Bobzien, M., Petzold, I., Burghardt, D. 2007, Automated Derivation of a 1:300 000 topographic map from Swiss DLM VECTOR 200, in: Proceedings of 23rd International Cartographic Conference. Moscow, Russia, CD-ROM.

Burghardt, D., Duchene, C. , Mackaness, W. (Eds), 2013, Abstracting geographic information in a data rich world: methodologies and applications of map generalisation. To be published by Springer.

Grünreich, D., 1985, Computer assisted generalisation: Papers CERCO-Cartography Course. Frankfurt am Main, Institut für Angewandte Geodasie.

Mackaness W., Ruas, A., Sarjakoski, T., (Eds) 2007, Generalisation of Geographic Information: Cartographic Modeling and Applications. Elsevier.

Revell, P., Regnauld, N., Bulbrooke,G., 2011, OS VectorMapTM District: Automated Generalisation, Text Placement and Conflation in Support of Making Pubic Data Public, in: Proceedings of 25rd International Cartographic Conference. Paris, France

Thom, S., 2005, A Strategy for Collapsing OS Integrated Transport Network™ dual carriageways. 8th ICA WORKSHOP on Generalisation and Multiple Representation, A Coruña, July 7-8th, 2005

Trevisan, J. 2004: From DLM to multi representation DCM - Modelling an application on buildings, ICA worshop on generalisation and multiple representations; 20-21 August 2004 – Leicester, UK.